Supplementary Information

Towards understanding the crystallization of photosystem II: Influence of poly(ethylene glycol) of various molecular sizes on the micelle formation of alkyl maltosides

Frank Müh,[1] Adrian Bothe,[2] Athina Zouni[3]

[1] Institut für Theoretische Physik, Johannes Kepler Universität Linz, Altenberger Strasse 69, A-4040 Linz, Austria.

[2] Institut für Molekularbiologie und Biophysik, ETH Zürich, HPK, Otto-Stern-Weg 5, CH-8093 Zürich, Switzerland.

[3] Institut für Biologie, Humboldt Universität zu Berlin, Leonor-Michaelis-Haus, Philippstrasse 13, D-10095 Berlin, Germany.

Corresponding author: Frank Müh, frank.mueh@jku.at

Content

Table S1: Original CMC values in molarity units (mM) obtained from graphical extrapolation of ANS titration curves.

M_r	% (w/v) PEG	CMC / mmol L^{-1}		
		DM	UDM	DDM
400	0	1.300*	0.380*	0.090*
	5	1.410	0.410	0.143
	10	1.770	0.495	0.207
	15	2.210	0.540	0.260
	20	3.030	0.685	0.430
	25	4.100	0.790	0.570
	30	5.200	–	–
550	0	1.300*	0.380*	0.090*
	2	1.400	0.390	0.094
	4	1.420	0.480	0.132
	8	1.650	0.570	0.132
	12	1.700	0.730	0.190
2000	0	1.300*	0.380*, 0.390	0.090*
	2	–	0.400*, 0.480	0.095*
	4	1.400*	0.450*, 0.570	0.110*
	5	1.290	0.430	0.147
	6	–	0.490*, 0.560	0.120*
	7.5	–	0.460	–
	8	1.750*	0.550*, 0.710	0.140*
	10	1.850	0.600*, 0.490, 0.740	0.140*, 0.224
	12	2.000*	0.650*, 0.740	0.180*
	12.5	–	0.805	–
	14	–	0.700*, 0.770	–
	15	2.640	0.880	0.305
	16	2.400*	0.750*, 0.900	0.180*
	20	3.300	–	–
8000	0	1.300*	0.380*	0.090*
	2	1.400		–
	4	1.600	0.480	0.130
	6	2.000		–
	8	–	0.560	0.190
	12	2.350	0.760	0.210
	16	2.400	–	–

* Published before: Müh F, DiFiore D and Zouni A (2015) The influence of poly(ethylene glycol) on the micelle formation of alkyl maltosides used in membrane protein crystallization. Phys Chem Chem Phys 17: 11678-11691.

Table S2 Results of linear regressions based on Eq. (52) performed separately for each PEG type. The fitted lines can be seen in Fig. S1-S3.

M_r	DM		UDM		DDM	
	slope	R^2	slope	R^2	slope	R^2
400	10.7 ± 0.3	0.9879	7.8 ± 0.2	0.9907	18.1 ± 0.7	0.9730
550	7.6 ± 0.5	0.9553	14.0 ± 0.7	0.9809	15.6 ± 1.6	0.9119
2000	10.5 ± 0.6	0.9337	12.7 ± 0.6	0.8250	15.4 ± 1.4	0.7758
8000	11.6 ± 1.1	0.8805	14.1 ± 0.5	0.9905	19.3 ± 1.7	0.9343

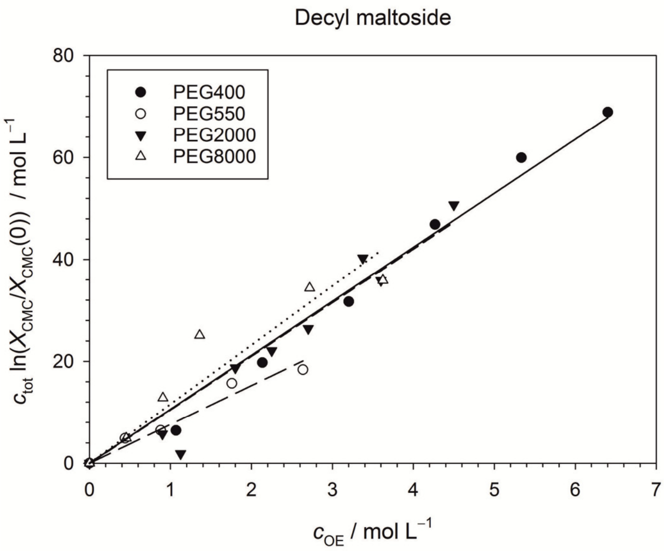

Fig. S1 Plots of $c_{\mathrm{tot}}\ \ln(X_{\mathrm{CMC}}(c_{\mathrm{OE}})/X_{\mathrm{CMC}}(0))$ (left side of Eq. (52)) versus c_{OE} for DM and the indicated PEG types. The slope of the plot is κ/p. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S2: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Note that the lines for PEG400 and PEG2000 are on top of each other. Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

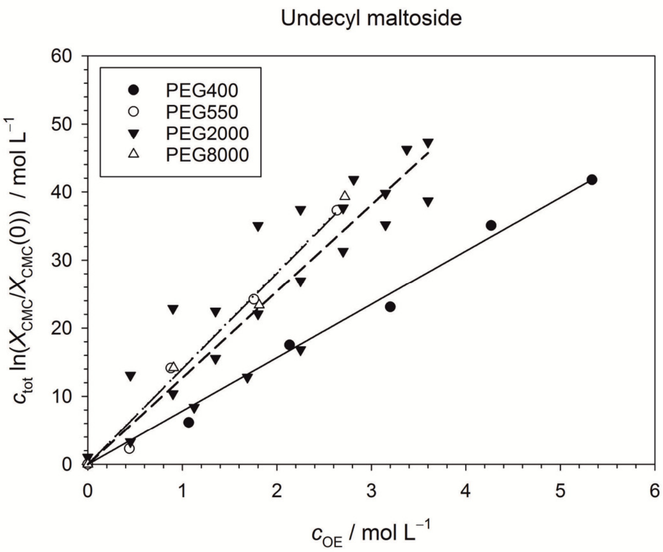

Fig. S2 Plots of $c_{\mathrm{tot}}\ \ln(X_{\mathrm{CMC}}(c_{\mathrm{OE}})/X_{\mathrm{CMC}}(0))$ (left side of Eq. (52)) versus c_{OE} for UDM and the indicated PEG types. The slope of the plot is κ/p. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S2: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Note that the lines for PEG550 and PEG8000 are on top of each other. Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

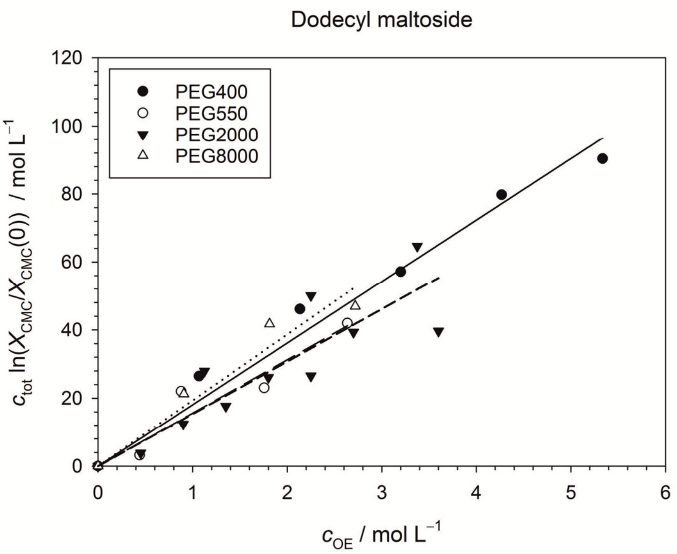

Fig. S3 Plots of $c_{\mathrm{tot}}\ \ln(X_{\mathrm{CMC}}(c_{\mathrm{OE}})/X_{\mathrm{CMC}}(0))$ (left side of Eq. (52)) versus c_{OE} for DDM and the indicated PEG types. The slope of the plot is κ/p. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S2: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Note that the lines for PEG550 and PEG2000 are on top of each other. Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

Table S3 Results of linear regressions based on plots of ln(CMC/CMC(0)) versus χ performed separately for each PEG type. The fitted lines can be seen in Fig. S4-S6.

M_r	DM		UDM		DDM	
	slope	R^2	slope	R^2	slope	R^2
400	0.0434 ± 0.0021	0.9645	0.0279 ± 0.0013	0.9732	0.0756 ± 0.0018	0.9901
550	0.0247 ± 0.0019	0.9395	0.0529 ± 0.0029	0.9749	0.0599 ± 0.0063	0.9012
2000	0.0410 ± 0.0029	0.9025	0.0501 ± 0.0026	0.7987	0.0617 ± 0.0059	0.7466
8000	0.0449 ± 0.0042	0.8821	0.0552 ± 0.0024	0.9842	0.0786 ± 0.0062	0.9419

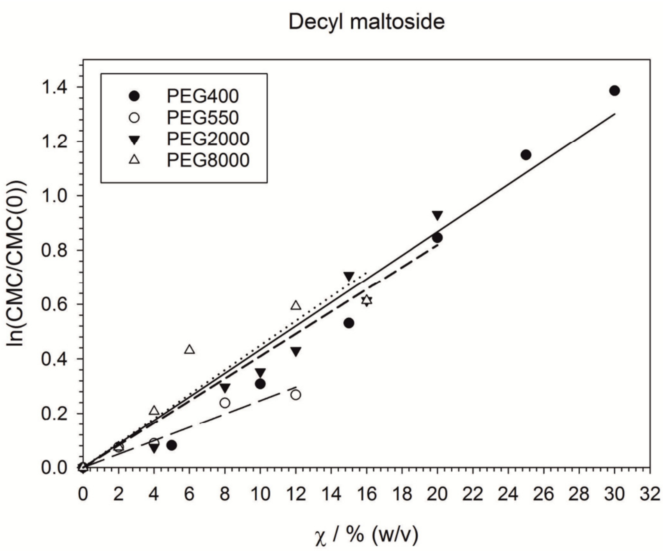

Fig. S4 Plots of ln(CMC/CMC(0)) versus χ for DM and the indicated PEG types. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S3: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

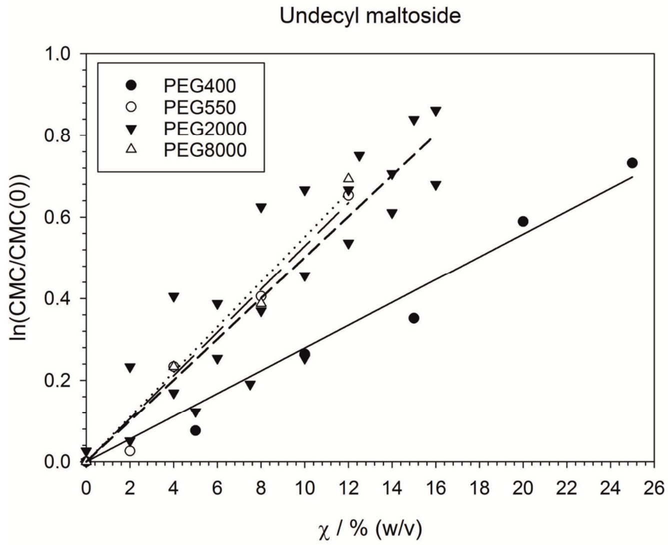

Fig. S5 Plots of ln(CMC/CMC(0)) versus χ for UDM and the indicated PEG types. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S3: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

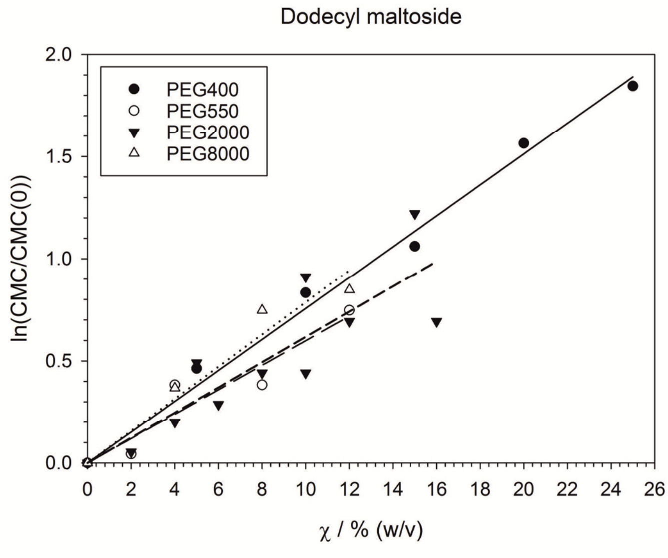

Fig. S6 Plots of ln(CMC/CMC(0)) versus χ for DDM and the indicated PEG types. The solid lines are linear regressions performed separately for each PEG type with zero intercept and the slopes given in Table S3: PEG400 (solid line), PEG550 (long dashed line), PEG2000 (short dashed line), PEG8000 (dotted line). Figure and fits made with SigmaPlot 13 (© 2014 Systat Software Inc.)

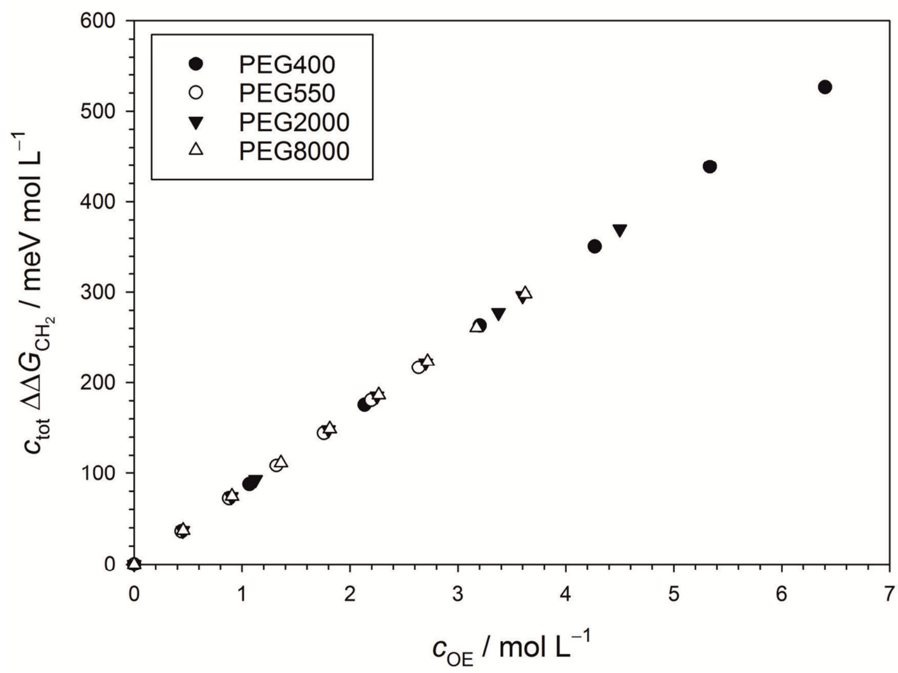

Fig. S7 Dependence of $c_{\mathrm{tot}}\Delta\Delta G_{\mathrm{CH_2}}$ on the concentration of OE-units, c_{OE}, in aqueous 100 mM PIPES (pH 7.0), 5 mM $CaCl_2$ computed on the basis of Eq. (59) for the PEG concentrations of the DM data sets (cf. SI Table S1) and $T = 298$ K. Figure made with SigmaPlot 13 (© 2014 Systat Software Inc.)

Derivation and Discussion of Eq. (36)

Eq. (36) is obtained by plugging Eq. (35) into Eq. (21) and using the limit $m \to \infty$. One may wonder whether it is justified to use Eq. (35) as an approximation for y in this context. To see this, we first estimate the order of magnitude of the terms occurring in Eq. (27). This estimate is based on the results of applying Eq. (36), so that it is rather a consistency check. The terms we want to evaluate contain JX_P, whose magnitude is given by

$$|JX_\mathrm{P}| = \frac{\kappa}{p} p \frac{c_\mathrm{P}}{c_\mathrm{tot}} = \frac{\kappa}{p} \frac{c_\mathrm{OE}}{c_\mathrm{tot}} = \frac{\kappa}{p} \frac{1}{c_\mathrm{tot}} \frac{10\,p}{M_\mathrm{P}} \chi < \frac{\kappa}{p} \frac{0.23\,\chi\,\mathrm{mol\,g^{-1}}}{c_\mathrm{tot}} \tag{S1}$$

To estimate an upper limit, we note that we have $\kappa/p < 20$ from the fits, and the experiments are done for $\chi < 50\ \%$ (w/v). For these values, we obtain $|JX_\mathrm{P}| = 7.5$. As an upper limit for y, we use $y = 2 \times 10^{-4}$ for DM as well as $y = 2 \times 10^{-5}$ for UDM and DDM based on the highest CMC values (using the fact that $x \approx y$ around the CMC). Using these data and $m = 85, 106, 140$ (Bothe et al. 2023), one can see that all terms to higher order in y in Eq. (27) can be neglected compared to terms linear in y or not containing y, which is true essentially because $y \ll 1$. Then, we have the approximation

$$x = y + \frac{m(m-2)y}{m^2(2m-1) + b_1 y} \quad \Longrightarrow \quad x = \frac{m^2(2m-1)y + b_1 y^2 + m(m-2)y}{m^2(2m-1) + b_1 y} \tag{S2}$$

Neglecting again terms quadratic in y, we obtain

$$x = \frac{m^2(2m-1)y + m(m-2)y}{m^2(2m-1) + b_1 y}$$

$$\Longrightarrow \quad m^2(2m-1)x + b_1 xy = m^2(2m-1)y + m(m-2)y \tag{S3}$$

Since x is in the same order of magnitude as y around the CMC, we can also neglect terms containing xy, so that b_1 drops out and we obtain Eq. (35). Plugging it into Eq. (21) yields

$$g_\mathrm{mic} = \frac{(m-1)}{m} \ln X_\mathrm{CMC} + \frac{1}{m} \ln\left\{\frac{m(2m^2-m)^m}{(m-2)(2m^2-2)^{m-1}}\right\} - JX_\mathrm{P}\left[X_\mathrm{CMC} \frac{(m-1)(2m^2-m)}{m(2m^2-2)} - 1\right] \tag{S4}$$

which for $J = 0$ coincides with the result found by Bothe et al. (2023). For $J \neq 0$, it contains an additional term. Eq. (36) is obtained in the limit $m \to \infty$ (with the definition of κ in Eq. (37)). Bothe et al. (2023) found this limit to be appropriate for DM, UDM, and DDM, and it is used here to facilitate the analysis. To further substantiate it, we show in Fig. S8 plots of the various functions of m occurring in Eq. (S4). It can be seen that the factor multiplying X_CMC in the last term on the right hand side of Eq. (S4) approaches unity as fast as the prefactor of $\ln X_\mathrm{CMC}$, so that both factors are of minor importance for aggregation numbers around 100. The logarithmic factor, which for $m \approx 100$ is in the order of 0.1 and not necessarily negligible, drops out in the computation of $\ln(X_\mathrm{CMC}/X_\mathrm{CMC}(0))$. Therefore, Eq. (36) is sufficient for an analysis of the PEG-induced CMC shift and still compatible with the refined definition of the CMC based on Eq. (22). Note, however, that we assume that m is not affected by PEG.

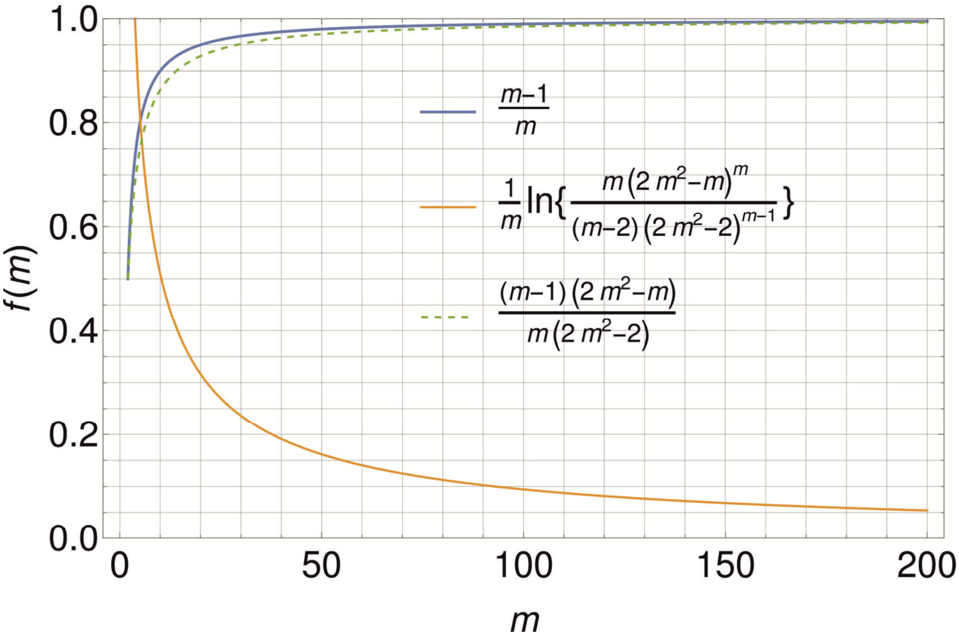

Fig. S8 Illustration of the various functions of m occurring in Eq. (S4). Plot and Figure made with Mathematica 13.0.1.0 (© 1988−2022 Wolfram Research)

Derivation and Discussion of Eq. (51)

We start by writing Eq. (48) in units of g dm^{-3} (at the risk of stating the obvious, we note that $\mathrm{L} = \mathrm{dm}^3$):

$$\rho = 1013.5\ \mathrm{g\,dm^{-3}} + 6.7\ \frac{\mathrm{g\,dm^{-3}}}{\mathrm{mol\,L^{-1}}}\ c_{\mathrm{OE}} \tag{S5}$$

Then, we plug Eq. (44) into Eq. (45):

$$c_{\mathrm{tot}} = \frac{\rho}{M_{\mathrm{wat}}} + c_{\mathrm{P}}\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right) + k_{\cos} = \frac{\rho}{M_{\mathrm{wat}}} + \frac{c_{\mathrm{OE}}}{p}\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right) + k_{\cos} \tag{S6}$$

Using Eq. (41) and the fact that $M_{\mathrm{wat}} = M_{\mathrm{H_2O}}$, we obtain

$$c_{\mathrm{tot}} = \frac{\rho}{M_{\mathrm{wat}}} + c_{\mathrm{OE}}\frac{\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right)}{\left(\frac{M_{\mathrm{P}} - M_{\mathrm{H_2O}}}{M_{\mathrm{OE}}}\right)} + k_{\cos} = \frac{\rho}{M_{\mathrm{wat}}} + c_{\mathrm{OE}}\frac{M_{\mathrm{OE}}}{M_{\mathrm{H_2O}}}\frac{\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right)}{\left(\frac{M_{\mathrm{P}}}{M_{\mathrm{H_2O}}} - 1\right)} + k_{\cos}$$

$$= \frac{\rho}{M_{\mathrm{wat}}} - c_{\mathrm{OE}}\frac{M_{\mathrm{OE}}}{M_{\mathrm{wat}}}\frac{\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right)}{\left(1 - \frac{M_{\mathrm{P}}}{M_{\mathrm{wat}}}\right)} + k_{\cos} = \frac{\rho}{M_{\mathrm{wat}}} - c_{\mathrm{OE}}\frac{M_{\mathrm{OE}}}{M_{\mathrm{wat}}} + k_{\cos} \tag{S7}$$

Plugging Eq. (S1) and the numerical values for the molar masses and $k_{\cos}$ into Eq. (S3), we finally obtain

$$c_{\mathrm{tot}} = \frac{1013.5\ \mathrm{g\,dm^{-3}} + 6.7\ \frac{\mathrm{g\,dm^{-3}}}{\mathrm{mol\,L^{-1}}}\ c_{\mathrm{OE}}}{18.015\ \mathrm{g\,mol^{-1}}} - c_{\mathrm{OE}}\frac{44.053\ \mathrm{g\,mol^{-1}}}{18.015\ \mathrm{g\,mol^{-1}}} - 1.5935\frac{\mathrm{mol}}{\mathrm{L}}$$

$$= 54.665\frac{\mathrm{mol}}{\mathrm{L}} - \left(\frac{6.7 - 44.053}{18.015}\right)c_{\mathrm{OE}} = 54.665\frac{\mathrm{mol}}{\mathrm{L}} - 2.073 c_{\mathrm{OE}} \tag{S8}$$

The somewhat counterintuitive result, that the number of moles for a given concentration of OE-units is independent of the actual number of OE-units in the polymer chain p, can be rationalized as follows: Let us assume for simplicity that we have only water and PEG. Let us further assume that one OE-unit displaces q water molecules from the solution and the end groups of PEG (H-…-OH) displace one water molecule. Then, $qp + 1$ moles are removed per one mole of PEG and one mole (i. e., the PEG chain itself) is added. Thus, if c_{wat} is the molarity of water, we have for the PEG solution

$$c_{\mathrm{tot}} = c_{\mathrm{wat}} - (qp + 1)c_{\mathrm{P}} + c_{\mathrm{P}} = c_{\mathrm{wat}} - qpc_{\mathrm{P}} = c_{\mathrm{wat}} - qc_{\mathrm{OE}} \tag{S9}$$

The assumption that the end groups displace one water molecule may be supported by the observation that the ratio of end groups to OE-units has no measurable influence on the density. For our buffer, we found $q = 2.073$. This result indicates that an OE-unit displaces about two water molecules, which have a smaller relative molecular weight (36.03) than one OE-unit (44.053) in accordance with the increase of the density of the PEG solution with PEG concentration. More precisely, $44.053 - 2.073 \times 18.015 = 6.708$ corresponds to the slope in Eq. (S5).